

(19) BUNDESREPUBLIK

DEUTSCHLAND



DEUTSCHES
PATENTAMT

(12) Offenlegungsschrift

(10) DE 44 25 527 A 1

(51) Int. Cl. 6:

~~C 12 N 15/74~~
~~C 12 N 15/75~~
~~C 12 N 15/70~~
~~C 12 N 15/31~~
~~C 12 N 1/21~~
~~C 07 C 227/18~~
~~C 07 K 14/195~~
~~// (C12N 1/21, C12R~~
~~1:07) (C12N 1/21,~~
~~C12R 1:19)~~

(71) Anmelder:

Vogelbusch Gesellschaft m.b.H., Wien, AT

(74) Vertreter:

H. Weickmann und Kollegen, 81679 München

(72) Erfinder:

Lubitz, Werner, Dr., Wien, AT

(54) S-Layer-Signalsequenz

(57) Die vorliegende Erfindung betrifft eine Nukleinsäure, die für ein Leader- bzw. Signalpeptid des S-Layerproteins von *B. stearothermophilus* kodiert und die gegebenenfalls in operativer Verknüpfung mit einer 3'-seitig angeordneten Protein-kodierenden Nukleinsäure oder/und 5'-seitig angeordneten Expressionskontrollsequenz ist.

DE 44 25 527 A 1

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

BUNDESDRUCKEREI 11.95 508 064/147

9/35

DE 44 25 527 A 1

Beschreibung

Die vorliegende Erfindung betrifft eine Nukleinsäure, die für ein Leader- bzw. Signalpeptid des S-Layerproteins von *B. stearothermophilus* kodiert und die gegebenenfalls in operativer Verknüpfung mit einer 3'-seitig angeordneten Protein-kodierenden Nukleinsäure oder/und 5'-seitig angeordneten Expressionskontrollsequenz ist.

Kristalline bakterielle Zelloberflächenlayer (S-Layer) bilden in vielen Eubakterien und den allermeisten Archaeabakterien, die äußerste Zellwandkomponente (Sleytr et al. (1988), Crystalline Bacterial Cell Surface Layers, Springer Verlag Berlin; Messner und Sleytr (1992), Adv. Microb. Physiol. 33, 213–275). Die meisten der gegenwärtig bekannten S-Layerproteine sind aus identischen Glykoproteinen zusammengesetzt, die scheinbare Molekulargewichte im Bereich von 40.000 bis 220.000 aufweisen. Die Komponenten von S-Layern sind selbst-assemblierend und die meisten Gitter haben eine schräge (p2), quadratische (p4) oder hexagonale (p6) Symmetrie. Die Funktionen von bakteriellen S-Layern sind immer noch nicht vollständig bekannt, aber aufgrund ihrer Lokalisierung an der Oberfläche dürften die porösen kristallinen S-Layer hauptsächlich als Schutzhüllen, Molekularsiebe oder zur Förderung der Zelladhäsion und Oberflächenerkennung dienen.

Genetische Daten und Sequenzinformationen sind für verschiedene S-Layergene aus Mikroorganismen bekannt. Eine Übersicht findet sich bei Peyret et al. (1993), Mol. Microbiol. 9, 97–109. Auf diese Daten wird ausdrücklich Bezug genommen. In dieser Arbeit ist die Sequenz des für das S-Layer-Protein von *B. stearothermophilus* PV72 kodierenden Gens und ein Verfahren zu dessen Klonierung angegeben.

B. stearothermophilus PV72 ist ein gram-positives Bakterium, das mit einem hexagonal angeordneten S-Layer bedeckt ist. Die Hauptkomponente des *B. stearothermophilus* S-Layer ist ein 128 kDa Protein, bei dem es sich um das häufigste Protein der Zelle mit einem Anteil von ungefähr 5% bezüglich der gesamten Proteinbestandteile handelt. Es sind verschiedene Stämme von *B. stearothermophilus* charakterisiert worden, die hinsichtlich des Typs von S-Layer Gitter, dem Molekulargewicht und der Glykosilierung der S-Layer-Komponenten unterschiedlich sind (Messner und Sleytr (1992), Supra). Um die Regulation von Synthese, Transport und Variabilität von S-Layerproteinen zu verstehen, besteht ein starkes Bedürfnis daran, die Primärstruktur dieses Proteins inklusive seines Signal- bzw. Leaderpeptids zu bestimmen.

Ein Gegenstand der Erfindung ist somit eine Nukleinsäure, welche dadurch gekennzeichnet ist, daß sie für ein funktionelles Signalpeptid kodiert und

- 30 (a) den Signalpeptid-kodierenden Abschnitt, der in SEQ. ID. NO. 1 dargestellten Nukleotidsequenz,
- (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
- (c) eine zu den Sequenzen aus (a) oder/und (b) mindestens 90% homologe Nukleotidsequenz aufweist.

35 In SEQ. ID. NO. 1 ist die kodierende Nukleotidsequenz des S-Layergens aus *B. stearothermophilus* einschließlich des Signalpeptid-kodierenden Abschnitts und der davon abgeleiteten Aminosäuresequenz, gezeigt. Weiterhin sind auch 36 Nukleotide aus dem 5'-nicht-kodierenden Bereich und 112 Nukleotide aus dem 3'-nicht-kodierenden Bereich gezeigt.

40 Das S-Layergen von *B. stearothermophilus* kodiert für ein Protein mit insgesamt 1229 Aminosäuren einschließlich eines N-terminalen Signalpeptids mit 30 Aminosäuren. Die Spaltstelle zwischen dem Signalpeptid und reifem Protein befindet sich zwischen Position 30 und 31 der Aminosäuresequenz. Das Signalpeptid des in SEQ. ID. NO. 1 gezeigten S-Layerproteins weist eine basische aminoterminale Domäne gefolgt von einer hydrophoben Domäne auf.

45 Sequenzvergleiche mit anderen Signalpeptiden zeigten eine gewisse Homologie zu Signalpeptiden von extrazellulären Proteinen in Bazillen, wie etwa alkalische Phosphatase und neutrale Phosphatase von *B. amyloliquefaciens* (Vasantha et al. (1984), J. Bacteriol. 159, 811–819) sowie mit den Signalpeptiden für das *B. sphaericus* Gen 125 (Bowditch et al. (1989), J. Bacteriol. 171, 4178–4188) und das OWP-Gen von *B. brevis* (Tsuboi et al. (1986), J. Bacteriol. 168, 365–373).

50 Die erfindungsgemäße Signalpeptid-kodierende Nukleinsäure kann in operativer Verknüpfung mit einer 3'-seitig angeordneten, Protein-kodierenden Nukleinsäure sein, so daß die für das Signalpeptid kodierende Nukleinsäure und die Protein-kodierende Nukleinsäure eine genetische Fusion bilden, deren Genprodukt ein Polypeptid ist, das ein Signalpeptid und einen reifen Proteinanteil enthält.

55 Das von der Protein-kodierenden Nukleinsäure kodierte Protein kann ein eukaryontisches oder prokaryontisches Protein, z. B. das homologe S-Layerprotein, ein anderes S-Layerprotein oder jedes andere beliebige heterologe Protein sein.

Insbesondere kann die Protein-kodierende Nukleinsäure

- 60 (a) den Protein-kodierenden Abschnitt der in SEQ. ID. NO. 1 dargestellten Nukleotidsequenz,
- (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
- (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierende Nukleotidsequenz aufweisen.

65 Unter dem Begriff "stringente Hybridisierung" im Sinne der vorliegenden Erfindung versteht man, daß eine Hybridisierung auch nach Waschen bei 55°C, vorzugsweise 60°C in einem wäßrigen Niedrigsalz-Puffer (z. B. 0,2 × SSC) noch auftritt (siehe auch Sambrook et al. (1989), Molecular Cloning. A Laboratory Manual).

Weiterhin kann das von der Protein-kodierenden Nukleinsäure kodierte Protein auch ein künstliches Poly-

peptid sein, das einen hohen Gehalt an seltenen Aminosäuren (z. B. Met, Trp und insbesondere Lys) enthält. Die Konstruktion derartiger künstlicher Polypeptide orientiert sich vorzugsweise an Teilsequenzen des nativen Proteins, um eine bessere Sekretion zu gewährleisten.

Besonders bevorzugt weist das Protein einen Lysin-Gehalt von mindestens 10% auf Basis der gesamten Aminosäuren auf. Einen derartigen Lysingehalt zeigt z. B. das in SEQ. ID. NO. 1 dargestellte native S-Layerprotein aus *B. stearothermophilus*.

In einem weiteren Aspekt der vorliegenden Erfindung kann die Signalpeptid-kodierende Nukleinsäure in operativer Verknüpfung mit einer 5'-seitig angeordneten funktionellen Expressionskontrollsequenz sein, d. h. mit einer Nukleinsäure, welche Transkriptions- und Translations-Regulationssignale enthält. Die Expressionskontrollsequenz kann an sich beliebig sein, sie wird zweckmäßigerweise anhand des zur Expression der Nukleinsäure verwendeten Wirtsorganismus und den Kultivierungsbedingungen ausgewählt. Einem Fachmann auf dem Gebiet der Molekularbiologie sind zahlreiche Beispiele für geeignete eukaryontische und prokaryontische Expressionskontrollsequenzen bekannt, so daß hier im einzelnen nicht darauf eingegangen werden muß.

In einer bevorzugten Ausführungsform der vorliegenden Erfindung steht die Signalpeptid-kodierende Nukleinsäure jedoch in operativer Verknüpfung mit der nativen Expressionskontrollsequenz des *B. stearothermophilus*-S-Layergens oder einer funktionellen Variante davon, die insbesondere

- (a) mindestens einen Abschnitt der in SEQ. ID. NO. 2 dargestellten Nukleotidsequenz,
- (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechenden Nukleotidsequenz oder
- (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierenden Nukleotidsequenz aufweist.

Besonders bevorzugt weist die Expressionskontrollsequenz eine Homologie von mindestens 80%, vorzugsweise mindestens 90% zu der in SEQ. ID. NO. 2 dargestellten Expressionskontrollsequenz auf.

In SEQ. ID. NO. 2 ist der 5'-nicht-translatierte Bereich des *B. stearothermophilus* S-Layergens in einer Länge von 247 Nukleotiden gezeigt, der die Expressionskontrollsequenz umfaßt.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Polypeptid, das von einer erfindungsgemäßen Nukleinsäure kodiert ist, d. h. ein funktionelles Signalpeptid gegebenenfalls in operativer Verknüpfung mit einer am Carboxyterminus des Signalpeptids angeordneten Proteinsequenz.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein rekombinanter Vektor, der mindestens eine Kopie einer erfindungsgemäßen Nukleinsäure enthält. Der Vektor kann in Eukaryonten, Prokaryonten oder in Eukaryonten und Prokaryonten replizierbar sein. Er kann ein in das Genom der Wirtszelle integrierbarer Vektor (z. B. Bacteriophage Lambda) oder ein Vektor sein, der extrachromosomal vorliegt (z. B. ein Plasmid). Vorzugsweise ist der erfindungsgemäße Vektor ein Plasmid, insbesondere ein prokaryontisches Plasmid.

Noch ein weiterer Gegenstand der vorliegenden Erfindung ist eine Wirtszelle, die mit einer Nukleinsäure oder einem rekombinanten Vektor gemäß vorliegender Erfindung transformiert ist. Die Zelle kann sowohl eine eukaryontische als auch eine prokaryontische Zelle sein. Vorzugsweise ist die Zelle ein prokaryontischer Organismus, besonders bevorzugt ein gram-positiver prokaryontischer Organismus und am meisten bevorzugt ein Organismus der Gattung *Bacillus*. Verfahren zur Transformation von eukaryontischen und prokaryontischen Zellen mit Nukleinsäuren sind allgemeiner Stand der Technik und brauchen daher nicht zu erläutert werden.

Noch ein weiterer Gegenstand der vorliegenden Erfindung ist ein Verfahren zur rekombinanten Herstellung von Proteinen, welches dadurch gekennzeichnet ist, daß man

- (a) eine Wirtszelle bereitstellt, die mit einer erfindungsgemäßen Signalpeptid-kodierenden Nukleinsäure in operativer Verknüpfung mit einer Protein-kodierenden Nukleinsäure transformiert ist,
- (b) die Wirtszelle unter solchen Bedingungen kultiviert, die mit einer Expression der Nukleinsäure und zu einer Erzeugung und Sekretion des davon kodierten Polypeptids führen und
- (c) das resultierende Polypeptid aus dem Kulturmedium gewinnt.

Als Wirtszelle für dieses Verfahren verwendet man vorzugsweise einen prokaryontischen Organismus, insbesondere einen Organismus aus der Gattung *Bacillus* und am meisten bevorzugt den Organismus *B. stearothermophilus*. Der Organismus bevorzugt ein stark zuckerhaltiges, osmotisch starkes Medium und hat den Vorteil einer hohen Produktionsrate und hohen Züchtungstemperaturen.

Die Expression von rekombinanten Polypeptiden unter Verwendung der erfindungsgemäßen Signalpeptid-kodierenden Nukleinsäure zeigt den Vorteil einer extrem hohen Effizienz aufgrund der guten Sekretionsleistung durch das erfindungsgemäße Signalpeptid und aufgrund der Tatsache, daß *B. stearothermophilus* ein Temperatur-resistenter Organismus mit einer optimalen Wachstumstemperatur im Bereich von 60°C ist, der eine schnelle Expression rekombinanter Polypeptide ermöglicht.

Zur verbesserten Gewinnung von Polypeptiden oder Proteinen aus dem Kulturmedium kann die Protein-kodierende Nukleinsäure einen Stretch geladener (z. B. Lys) oder hydrophober (z. B. Trp) Reste oder/und eine Streptavidin-Teilsequenz enthalten, die zur Bindung an Biotin in der Lage ist.

In einem Aspekt betrifft dieses erfindungsgemäße Verfahren die rekombinante Herstellung von S-Layerproteinen, insbesondere des 128 kDa S-Layer Proteins von *B. stearothermophilus*.

In einem weiteren Aspekt betrifft das erfindungsgemäße Verfahren die rekombinante Herstellung von Proteinen mit einem hohen Gehalt an essentiellen Aminosäuren, insbesondere einem Lysingehalt von mindestens 10% auf Basis der gesamten Aminosäuren. Zur Gewinnung dieser essentiellen Aminosäuren wird das erhaltene Protein vorzugsweise hydrolytisch (z. B. durch Säure oder/und durch Proteasen) gespalten und die bei der

Spaltung entstehenden Aminosäuren auf an sich bekannte Weise (z. B. durch chromatographische Methoden) isoliert. Beispiele für die Proteasespaltung von Proteinen sind bei Qi et al. (Plant. Physiol. 104 (1994), 127–133), Bransom und Dreher (Virology 198 (1994), 148–154) und Di Ianni et al. (J. Biol. Chem. 268 (1993), 2048–2051) angegeben. Auf diese Offenbarung wird hiermit Bezug genommen.

Schließlich betrifft die Erfindung auch noch eine Expressionskontrollsequenz, die (a) mindestens einen Abschnitt der in SEQ. ID. NO. 2 dargestellten Nukleotidsequenz, (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierende Nukleotidsequenz aufweist. Diese Expressionskontrollsequenz ermöglicht eine starke Expression, so daß z. B. das S-Layerprotein in einem Anteil von 5–10% des zellulären Gesamtproteins exprimiert wird.

Weiterhin wird die vorliegende Erfindung durch die folgenden Beispiele in Verbindung mit den Sequenzprotokollen SEQ. ID. NO. 1–6 und Fig. 1 erläutert.

Es zeigen:

SEQ. ID. NC. 1 die vollständige Nukleotidsequenz des kodierenden Abschnitts des S-Layergens von *B. stearothermophilus* und die davon abgeleitete Aminosäuresequenz,
 SEQ. ID. NO. 2 den 5'-nicht-translatierten Bereich des *B. stearothermophilus* S-Layergens,
 SEQ. ID. NO. 3 das Oligonukleotid P1,
 SEQ. ID. NO. 4 das Oligonukleotid P2,
 SEQ. ID. NO. 5 das Oligonukleotid P3,
 SEQ. ID. NO. 6 das Oligonukleotid P4 und

Fig. 1 eine schematische Darstellung der klonierten und durch PCR erhaltenen Genfragmente aus *B. stearothermophilus*.

Beispiel 1

Herstellung einer *B. stearothermophilus* PV72 Genbank

Eine im Expressionsvektor pSK(+) (Short et al., Nucleic Acids Res. 16 (1988), 7583–7600) durch partielle HindIII-Spaltung chromosomaler *B. stearothermophilus* pV72 DNA erzeugte Genbank wurde im E.coli Stamm JM83 (Vieira und Messing (1982), Gene 19, 259–268) hergestellt. Die Kolonien dieser Genbank wurden mit polyklonalen Antikörpern gegen gereinigtes S-Layerprotein gemustert. Durch die Antikörper-Musterung wurden drei immunreaktive Klone aus 10.000 rekombinanten Klonen identifiziert. In einer zweiten Musterungsrunde zeigte nur einer der drei Klone eine Kreuzreaktion mit S-Layer-spezifischen Antikörpern. Das in diesem Klon enthaltene Plasmid pBK24 enthielt eine HindIII-Insertion mit insgesamt ca. 2,3 kb Länge (Fig. 1, Konstrukt A).

Eine DNA-Sequenzierung der Insertion ergab einen offenen Leserahmen mit 2,3 kb einschließlich des 3'-Endes des S-Layergens.

Um eine vollständige Kopie des S-Layergens zu isolieren, wurde eine zweite Genbank hergestellt. Hierzu wurde die genomische DNA von *B. stearothermophilus* PV72 mit dem Restriktionsenzym PstI gespalten und eine genomische Southern Hybridisierungsanalyse bei 65°C (Sambrook et al. (1989), Supra) unter Verwendung von radioaktiv markierter DNA aus der Insertion des Plasmids pBK24 durchgeführt. Diese Hybridisierungsonde ergab nur ein positives Signal mit einer 4,6 kb Bande aus dem Southern Blot, die isoliert, in pUC18 kloniert und in den E.coli Stamm JM103 (Messing et al. (1981), Nucleic Acids Res. 9, 309–321) transformiert wurde. Von 2.000 Transformanten ergaben 12 rekombinante Klone positive Signale bei Musterung durch *in situ* Kolonie-Hybridisierung mit einem (³²P)-markierten, Sequenz-spezifischen Oligonukleotid P1 (SEQ. ID. NO. 3). Diese Klone wurden anschließend durch Restriktionsanalyse untersucht. Dabei wurde gefunden, daß sie identische 5'-Bereiche aufwiesen und innerhalb des offenen Leserahmens 150 bp stromaufwärts der HindIII Restriktionsstelle aus pBK24 endeten. Ein ähnlicher Versuch wurde unter Verwendung des Vektors pPLcat10 (Stanssens et al., Gene 36 (1985), 211–223) unternommen. Erneut wurde gefunden, daß alle positiven Klone an der gleichen Stelle innerhalb des offenen Leserahmens endeten.

Die Versuche zur Klonierung des vollständigen S-Layer Gens einschließlich der Signalsequenz in E.coli waren somit nicht erfolgreich.

Beispiel 2

Erzeugung der 5'-Region des S-Layer Gens von *B. stearothermophilus* PV72 durch PCR

Das erste PCR-Fragment (Fig. 1, Konstrukt B) wurde unter Verwendung der Primer P1 (SEQ. ID. NO. 3) und P2 (SEQ. ID. NO. 4) synthetisiert. Der Primer P2 enthält Sequenzen aus dem aminoterminalen Abschnitt des reifen S-Layergens und Restriktionsstellen. Die Sequenzinformation zur Herstellung dieses Primers wurde durch Bestimmung der N-terminalen Aminosäuresequenz des reifen S-Layerproteins erhalten. Hierzu wurde das gereinigte S-Layerprotein einer 2-D-Gelelektrophorese unterworfen, auf eine Polyvinylidendifluorid-Membran überführt und einer Mikrosequenzierung (Kalkkinen und Tilgmann (1988), J. Protein Chem. 7, 242–243) unterzogen.

Der Antisense-Primer P1 (SEQ. ID. NO. 3) enthält eine S-Layergen-spezifische Region (entsprechend nt 1465 bis nt 1482, 60 bp stromabwärts der 5'-HindIII-Stelle von pBK24) und Restriktionsstellen. Das unter Verwendung beider Primer erhaltene 1,3 kb PCR-Fragment wurde als PstI-Fragment unter Verwendung der flankierenden Restriktionsstellen in den Vector pUC18 kloniert und in den E.coli Stamm JM 103 transformiert.

Es wurde die DNA-Sequenz der PstI-Insertion bestimmt. Die resultierenden Sequenzdaten zeigten, daß

erneut eine 5'-Deletion stromaufwärts der HindIII-Stelle aufgetreten war.

Eine direkte Sequenzierung des PCR-Fragments I zeigte, daß dieses Fragment einen offenen Leserahmen (ORF) von 1,3 kb enthielt. Diese Ergebnisse bestätigten, daß die ORF-Sequenz des PCR-Fragments I die DNA-Sequenz des reifen S-Layerproteins von *B. stearothermophilus* PV72 vervollständigte.

Um weitere 5'-stromaufwärts gelegene Bereiche des sbsA-Gens zu erhalten, wurde das PCR-Fragment I als Hybridisierungssonde für eine Southern Blot Analyse einer BamHI-gespaltenen genomischen DNA aus *B. stearothermophilus* verwendet. Das entsprechende Fragment der hybridisierenden Bande (5,8 kb) wurde aus einem Agarose-Gel isoliert und mit BamHI-gespaltenem pUC18 nach Standardmethoden ligiert. Zur Bestimmung weiterer 5'-Sequenzen des SbsA-Gens wurde ein zweites PCR-Fragment (Fig. 1, Konstrukt C) unter Verwendung des pUC18-spezifischen Primers P3 (SEQ. ID. NO. 5) und des sbsA-Gen-spezifischen Primers P4 (SEQ. ID. NO. 6) hergestellt. Das entsprechende PCR-Fragment II hatte eine Größe von 0,7 kb. Eine direkte Sequenzierung des PCR-Fragments II zeigte, daß sich 5'-seitig des sbsA-ORF eine Signalsequenz von 30 Codons befand.

Beispiel 3

Sequenzanalyse des SbsA-Gens

Das S-Layer Gen von *B. stearothermophilus* weist einen offenen Leserahmen in einer Länge von 3.687 bp auf, der für ein Protein mit 1.228 Aminosäuren kodiert. Der ORF startet mit einem ATG an Nukleotidposition 1 (SEQ. ID. NO. 1), von dem 5'-seitig eine ribosomale Bindungsstelle (rbs) bei Position -12 angeordnet ist. Die durch Edman-Abbau des reifen Proteins und tryptischer Peptide davon bestimmten Aminosäuren sind unterstrichen. Direkte Nukleotid-Repeatsequenzen sind doppelt unterstrichen.

46 bp stromabwärts des TAA Stopcodons des sbsA-Gens wurde eine palindrome Sequenz gefunden, die als "Hairpin"-Sekundärstruktur zur Termination der Transkription durch die RNA-Polymerase dienen kann und durch Linien mit Pfeilen bezeichnet ist.

Die ersten 30 Aminosäuren des SbsA-Proteins bilden eine Signalsequenz. Die Spaltung des Signalpeptids findet nach einem Sequenzmotiv Ala-Ala statt (Pfeil), was häufig bei Signalpeptiden gefunden wird (Perlman und Halvorson (1983), J. Mol. Biol. 167, 391—409).

Die 5'-untranslatierte Sequenz des sbsA-Gens, mit der Expressionskontrollsequenz ist in SEQ. ID. NO. 2 gezeigt. Die rbs Region sowie die Bereiche bei -10 und -35 bezüglich des Transkriptionsstarts sind unterstrichen.

5

10

15

20

25

30

35

40

45

50

55

60

65

SEQ. ID. NO. 1

NAME: sbsA Gen und Protein aus B. stearothermophilus

ART DER SEQUENZ: Nukleotid und Aminosäure

LÄNGE: 3706 Nukleotide

1228 Aminosäuren

TCTAAAGTAAGTAAAMGAACTGGATATACTCTCA

S.D.

1 M D R K K A V K L A T A S A I A A S A F V A A N P N A S E A
 1 ATGGATAGGAAAAGCTGTGAACCTAGCCACGCAAGTGCTATTGAGCAACTGCATTGTCCTGCAATGCCAACGCTTGAAACG|

10 31 A T D V A T V V S O A K A O E K K A Y Y T Y S H T V T E T G
 91 GCTACAGAATGAGCAACAGTACTTAAGCCAAAGCAGCTTCAGGAAAGCATACTATACTTACAGCCATACAGTAAACGAAAGCTG

15 61 F F P N I N D V Y A E Y N K A K K R Y R D A V A L V N K A G
 181 GAATTCCCAACATTAAGGATGTATAGCTGAATACACAGGAAAGGATACCGTACCGGTTACCTAGTGAATTAACGCT

91 G A K K D A Y L A D I L O K F Y E T Y V F K A N P K S G E A R
 271 GCGCGAAAAAGACGGCTACTTACCTGATTTACAAAAGATAATGAAACTTACGGTTTCAAGCAGCCTAAATCTGGCGAACGCTGT

20 121 V A T Y I D A Y N Y A T K L D E M R Q E L E A A V Q A K D L
 361 GTAGCAACTTACATOGATGCCATACAATGCAACAAAATAGACGAAATGCGCAAGCTAGAGCTAGGGCTGCTGTTCAAGCAAAAGATTG

151 451 E K A E Q Y Y H X I P Y E I K T R T V I L D R V Y G K T T R
 GAAAAGCAGAACAACTATCATACTACAAAATCCTATGAATTAAACTCGCACAGTCATTAGATGCGTATATGCTAAACAACTGCT

25 181 D L L R S T F K A K A Q E L R D S L I Y D I T V A M K A R E
 541 GATTACTTGGCTCTACATTTAAAGCAAAGCACAAGAACTTGGGACAGCTTAAATATGATATTACCGTTGCAATGAAACGGCGGA

211 631 V Q D A V K A G N L D K A K A A V D Q I N Q Y L P ' K V T D A
 GTACAAGACGGCTGTGAAGCAGCAATTAGACAAAGCTAAAGCTGCTGCTGATCAATCAATCACTTACCAAAAGTAACAGATGCT

30 241 721 F K T E L T E V A K K A L D A D E A A L T P K V E S V S A I
 TTCAAAACTGAACTAACAGAAGTAGCGAAAAACATTAGATGCCAGATGAAAGCTGGCTACTCCAAAGTGAAGCTAGGGATT

271 811 N T Q N K A V F L T A V P V N G T L K L Q L S A A A N E D T
 AACACTCAAACAAACGCTGTGAATTCAGCAGTACCAAGGACACTAAATACAACTTACGGCTGCTGCAATGAAGATAACAA

35 301 901 V N V N T V R I Y K V D G N I P F A L N T A D V S L S T D G
 GTAAACGTAATACIGTAGTATCTAAAGTGGACGGTAAACATCCATTGCGCTAAACGGCAGATGTTCTTACAGACGGAA

331 991 K T I T V D A S T P F E N N T E Y K V V V V K G I K D K N G X
 AAAACTATCACTTGGATCTTCAACTCCATTGAAATAACCGAGTATAAGTAGTATGCTAAAGCTTAAAGACAAATGCGAA

40 361 1081 E F K E D A F T F K L R N D A V V T Q V F G T N V T N N T S
 GAATTAAAGAAGATGCATTCACTTCAAGCTTCAAGCTGAAATGATGCTCTAGTACTGAGTGTGGAACTAATGTAACAAACACTCT

391 1171 V N L A A G T F D T D D T L T V V F D K L L A P E T V N S S
 GTAAACCTAGCACCAGGTACTTTCGACACTGACGATACTTAAACAGTAGTATTGATAAGTGTGACCTGAAACTGAAACAGCTG

421 1261 N V T I T D V E T G K R I P V I A S T S G S T I T I T L K E
 AACCTTACATACAGATGTTGAAACTGGAAAGCGATTCAGTAATGCTACTCTGGCTCTACAAATCTTACAGTAAAGCTAAACAA

50 451 1351 A L V T G K Q Y K L A I N N V K T L T G Y N A E A Y E L V F
 CGGTTACTAACTGCTAAACATATAAACTTGCTATCAATATTTAAACATTAACGTTACATGCCAGAGCTTACGAGTTAGTGTTC

421 1441 T A N A S A P T V A T A P T T T G G T T L S T G S L T T N V
 ACTGCAACGCACTGCAACCAACTGCTGCTACOGCTCTACTACTTACGGTGTACAACCTTATCTACTGGCTCTTACAAACAA

55 511 1531 W G K L A G G V N E A G T Y Y P G L Q F T T T F A T K L D E
 TGGGCTAAATGGCTGGTGTGAATGAAAGCTGAAACTTATCTGGCTCTCAATCACAACAAAGCTTACGAGCTTACGAGCTAA

541 1621 S T L A D N F V L V E K E S G T V V A S E L K Y N A D A K M
 TCTACTTTAGCTGATAACTTGTATTGATGAAAAGAATCTGTTACAGTGTGCTCTGACTAAATATAATGCCAGGCTAAACG

60 571 1711 V T L V P K A D L K E N T I Y Q I K I K K G L K S D K G I E
 GTAACTTTAGTGCACAAAGGGACCTTAAGAAAATACATCTATCAATCAATTAAGGCTTGAAGCTGGATAGGTTAGCTAAAGCTAA

65 601 1801 L G T V N E K T Y E F K T Q D L T A P T V I S V T S K N G D
 TTAGGCACCTTAAAGGCAAAACATAGGTTCAAACTCAAGACTTAACTGCTCTACAGTTAACTGGTACCTCTAAAGCTGAC

631 1891 A G L K V T E A Q E F T V K F S E N L N T F N A T T V S G S
 CCTGGATTAAAGTAACGCTCAAGAACTTACTGTAAGTGTGAACTTCTCAGAGAAATTAAACATTTATGCTACAAACGTTGGCTAGC

661 T I T Y G Q V A V V K A G A N L S A L T A S D I I P A S V E
1981 A C A A T C A C A T A C O G T C A A G T T G C T G T A G T A A A G C G G G T C O C A A C T T A C T C G C T C T T A C A G C A C T G A C A T C A T T C C A G G T C A G T G T G T G A A
691 A V T G Q D G T Y K V K V A A N Q L E R N Q G Y K L V V F G
2071 C C C G T T A C T G T C A A G A T G A A C A T A C A A G T G A A A G T T G C T A C C O M A T T G A A C G T A C C L A O G T A C A A A T T A G T A G T G T C C G T
721 K G A T A P V K D A A N A N T L A T N Y I Y T F T T E G Q D
2161 A A A G G T C O C A A C A G C T C T G T T A A A G A T G C T C O C A A A T G C C A A A T C T T A G C A A C T A C T A T A T C T A C A T T C A C A C T G A A G G T C A A G C
751 V T A P T V T K V F K G D S L K D A D A V T T L T N V D A G
2251 C T A C A G C A C C A A C O G T T A C C A A A G T T A C T C A A A G G T G A T T C I T T A A A G A C G T G A T G C A G T T A C T A C A C T T A C G A A C G T T G A T G C A G G T
781 Q K F T I Q F S E E L K T S S G S L V G G K V T V E R K L T N
2341 C A A A A T T C A C T A T C O C A A T T A G C G A A C A A T T A A A A C T T C A G T G C T C T T A G T G C T G C A A A G T A A C T G T O G A A A T T A C C A A C
811 N G W V D A G T G T T V S V A P K T D A N G K V T A A V V T
2431 A A C G G A T G G G T A C A T G C T G T A C T C G G A A C A A C T G T A T C A G T T G C T C C T A A G A C A G A T G C A A T G G T A A A G T A A C A G C T G C T G T G G T I A C A
841 L T G L D N N D K D A K L R L V V D K S S T D G I A D V A G
2521 T T A C T G G T C T T G A C A A T A A C G A C A A A G A T G C G A A A T T G G G T C T G G T A C T A G A A T G T C T C T A C T C A T G G A A T T G C T G A T G T G A C T G G T
871 N V I K E K D I L I R Y N S W R H T V A S V K A A A D K D G
2611 A A I G T A A T T A A G G A A A A G A T A T T T A A T T C G T T A C A A C A G C T U G A G A C A C A C T G T A G C T C T G T G A A A G C T G C T G T G A C A A A G A T G G T
901 Q N A S A A F P T S T A I D T T K S L L V E F N E T D L A E
2701 C A A A C G C T T C T G C A T T C C A A C A A G C A C T G C A A T G A P A C A A C T A A G G C T T A T T G T G A A T C A A T G A A A C T G A T T A G C G G A A
931 V K P E N I V V K D A A G H A V A G T V T A L D G S T N K E
2791 G T T A A A C C T G A G A A C A T G T G T T A A A G A T G C G C A G G T A A T C G G T A C T C G G T A C T I G T A A C A G C T T A G A C G G T T C T A C A A A T T A A T T
961 V F T P S O E L K A G T V Y S V T I D G V R D K V G N T I S
2881 C T A T T C A C T C A A G A T T A A A G C T G T A C A S T T A C T C G T G A C A M T T G A C G G T G T G A G A G A T A A A G T A G G T A A C A C A A T C T C T
991 K Y I T S F K T V S A N P T L S S I S I A D G A V N V D R S
2971 A A A T A C A T T A C T C T G T C A A G A C T G T A T C T G G A A T C C A A C G T T A T C T C A A T C A C T C T T A A G A A G G C T G A C C G A A C T T C A T T A C T A
1021 K T I T I E F S D S V P N P T I T L K K A D G T S F T N Y T
3061 A A A C A T T A C A A T T G A A T T C A G G G A T T C A G T T C C A A A C C C A A C A A T C A C T C T T A A G A A G G C T G A C C G A A C T T C A T T A C T A
1051 L V N V N N E N K T Y K I V F H R G V T L D E F T Q Y E L A
3151 T T A G T A A T G T A A T T A T G J A A A T A A A C A T A C A A A A T T G T A T T C C A C A A G G T G T A A C A C T T G A G G A G G T T T A C T C A A T A T G A G T T A S C A
1081 V S K D F Q T G T D I D S K V T F I T G S V A T D E V K P A
3241 G T T T C A A A G A T T T C A A A C T G G T A C T G A T A T I G C A T C A C A G G T T C T G T G C T A C T G A C C G A A G T A A A A C C T G C T
1111 L W G V G S W N G T S Y T Q D A A A T R L R S V A D F V A E
3331 C T A G T A O G G S T T G G T T C A T G G A A T G G A A C A A G C T A T A C T C A G G A T G C T G C A G C A A C C G A C T T C G G T C T G T A G G T C A C T T C T G C C G G A G
1141 P V A L Q F S E G I D L T N A T V T V T N I T D D K T V E V
3421 C C A G T T G C C C T T C A A T T C T C A G A A G G T A T G A T T E A C C G A A T G C A A C T G T G A C A G T A A C A A A T T A C T I G T A G T A A A A C T G T G A A G T T
1171 I S K E S V D A D H D A G A T K E T L V I N T V T P L V L D
3511 A T T C A A A G A G A G T G T A G A G C G A G C C A T G A T G C A G G T G C T A C T A G G A G A C T T A G T A A T T A C A C A G T T A C T C C T T A G T A C T T G I T
1201 N S K T Y K I V V S G V K D A A G N V A D T I T F Y I K S T O P
3601 A A C A G C A A G A C T T A T A A G A T T G T G A A T G O G A T T A A A G A T G C A C C G A G T A A T G T G C A G A T A C T A T T A C A T T A T T A A G T A A T C I
GGGCTAGGTGTTGTCACCGCTCAAGGTGTCAAAAATAGTGAAGAAAGCTCTGCCGAGAGAAAATCTCTGCCGGGCTTTCTTTTGCTC

AATCTGTATCAAGCTT

DE 44 25 527 A1

SEQ. ID. NO. 2

NAME: Expressionskontrollsequenz des sbsA Gens
aus B. stearothermophilus

ART DER SEQUENZ: Nukleotid

LÄNGE: 247 Nukleotide

15

1 CTG aaa gtc ttt tgt taa tga aaa gta tgc gat ttc cga aaa tcg ata caa aat tct caa 31

20

61 aag gtt tat agt caa cgt gaa ata tag att aaz aga aac ata tat aaa cat acg tca tca 91

25

121 ttt tct att aat cat CTC tta tag gtt aca tag gat gat aaa gta aat ttt tgg gga ttt 151

30

181 gta agc aat gaa tct att aat atz cat tgc TCT aaa gta aTg taz aat gaa ggg gat atz 211

35

241
ctt ctc a

40

45

50

55

60

65

SEQ. ID. NO. 3:

NAME: P1

ART DER SEQUENZ: Nukleotid

LÄNGE: 33 Nukleotide

AATGGATCCC TGCAGAGTAG GAGCGGTAGC AAC

5

10

SEQ. ID. NO. 4:

NAME: P2

ART DER SEQUENZ: Nukleotid

15

LÄNGE: 33 Nukleotide

ATAGGATCCC TGCAGGCTAC AGATGTAGCA ACA

20

SEQ. ID. NO. 5:

NAME: P3

ART DER SEQUENZ: Nukleotid

25

LÄNGE: 17 Nukleotide

CAGGAAACAG CTATGAC

30

SEQ. ID. NO. 6:

NAME: P4

ART DER SEQUENZ: Nukleotid

35

LÄNGE: 18 Nukleotide

GGCTGTAACT ATAGTATG

40

Patentansprüche

1. Nukleinsäure, dadurch gekennzeichnet, daß sie für ein funktionelles Signalpeptid kodiert und
 - (a) den Signalpeptid-kodierenden Abschnitt der in SEQ. ID. NO. 1 dargestellten Nukleotidsequenz,
 - (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
 - (c) eine zu den Sequenzen aus (a) oder/und (b) mindestens 90% homologe Nukleotidsequenz aufweist.
2. Nukleinsäure nach Anspruch 1, dadurch gekennzeichnet, daß sie in operativer Verknüpfung mit einer 3'-seitig angeordneten, Protein-kodierenden Nukleinsäure ist.
3. Nukleinsäure nach Anspruch 2, dadurch gekennzeichnet, daß das von der Protein-kodierenden Nukleinsäure kodierte Protein ein S-Layerprotein ist.
4. Nukleinsäure nach Anspruch 2 oder 3, dadurch gekennzeichnet, daß die Protein-kodierende Nukleinsäure
 - (a) den Protein-kodierenden Abschnitt der in SEQ. ID. NO. 1 dargestellten Nukleotidsequenz,
 - (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
 - (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierende Nukleotidsequenz aufweist.
5. Nukleinsäure nach einem der Ansprüche 2—4, dadurch gekennzeichnet, daß das von der Protein-kodierenden Nukleinsäure kodierte Protein einen Lysin-Gehalt von mindestens 10% auf Basis der gesamten Aminosäuren aufweist.
6. Nukleinsäure nach einem der Ansprüche 1—5, dadurch gekennzeichnet, daß sie in operativer Verknüpfung mit einer 5'-seitig angeordneten funktionellen Expressionskontrollsequenz ist.
7. Nukleinsäure nach Anspruch 6, dadurch gekennzeichnet, daß die funktionelle Expressionskontrollsequenz
 - (a) mindestens einen Abschnitt der in SEQ. ID. NO. 2 dargestellten Nukleotidsequenz,
 - (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
 - (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierende Nu-

kleotidsequenz aufweist.

8. Polypeptid, dadurch gekennzeichnet, daß es von einer Nukleinsäure nach einem der Ansprüche 1—7 kodiert ist.
9. Rekombinanter Vektor, dadurch gekennzeichnet, daß er mindestens eine Kopie einer Nukleinsäure nach einem der Ansprüche 1—7 enthält.
10. Rekombinanter Vektor, dadurch gekennzeichnet, daß er ein Plasmid ist.
11. Wirtszelle, dadurch gekennzeichnet, daß sie mit einer Nukleinsäure nach einem der Ansprüche 1—7 oder einem rekombinanten Vektor nach Anspruch 9 oder 10 transformiert ist.
12. Wirtszelle nach Anspruch 11, dadurch gekennzeichnet, daß sie ein gram-positiver prokaryontischer Organismus ist.
13. Wirtszelle nach Anspruch 12, dadurch gekennzeichnet, daß sie ein Organismus der Gattung Bacillus ist.
14. Verfahren zur rekombinanten Herstellung von Proteinen, dadurch gekennzeichnet, daß man
 - (a) eine Wirtszelle bereitstellt, die mit einer Nukleinsäure nach Anspruch 2 transformiert ist,
 - (b) die Wirtszelle unter solchen Bedingungen kultiviert, die zu einer Expression der Nukleinsäure und zu einer Erzeugung und Sekretion des davon kodierten Polypeptids führen, und
 - (c) das resultierende Polypeptid aus dem Kulturmedium gewinnt.
15. Verfahren nach Anspruch 14, dadurch gekennzeichnet, daß man als Wirtszelle einen gram-positiven prokaryontischen Organismus verwendet.
16. Anspruch nach Anspruch 15, dadurch gekennzeichnet, daß man einen Organismus der Gattung Bacillus verwendet.
17. Verfahren nach einem der Ansprüche 14—16 zur rekombinanten Herstellung von S-Layer-Proteinen.
18. Verfahren nach einem der Ansprüche 14—17 zur rekombinanten Herstellung von Proteinen mit einem Lysin-Gehalt von mindestens 10% auf Basis der gesamten Aminosäuren.
19. Verfahren nach einem der Ansprüche 14—18, weiterhin umfassend die Schritte der hydrolytischen Spaltung des Proteins und der Gewinnung von bei der Spaltung entstehenden Aminosäuren.
20. Expressionskontrollsequenz, dadurch gekennzeichnet, daß sie
 - (a) mindestens einen Abschnitt der in SEQ. ID. NO. 2 dargestellten Nukleotidsequenz,
 - (b) eine der Sequenz aus (a) im Rahmen der Degeneration des genetischen Codes entsprechende Nukleotidsequenz oder
 - (c) eine mit den Sequenzen aus (a) oder/und (b) unter stringenten Bedingungen hybridisierende Nukleotidsequenz aufweist.

Hierzu 1 Seite(n) Zeichnungen

35

40

45

50

55

60

65

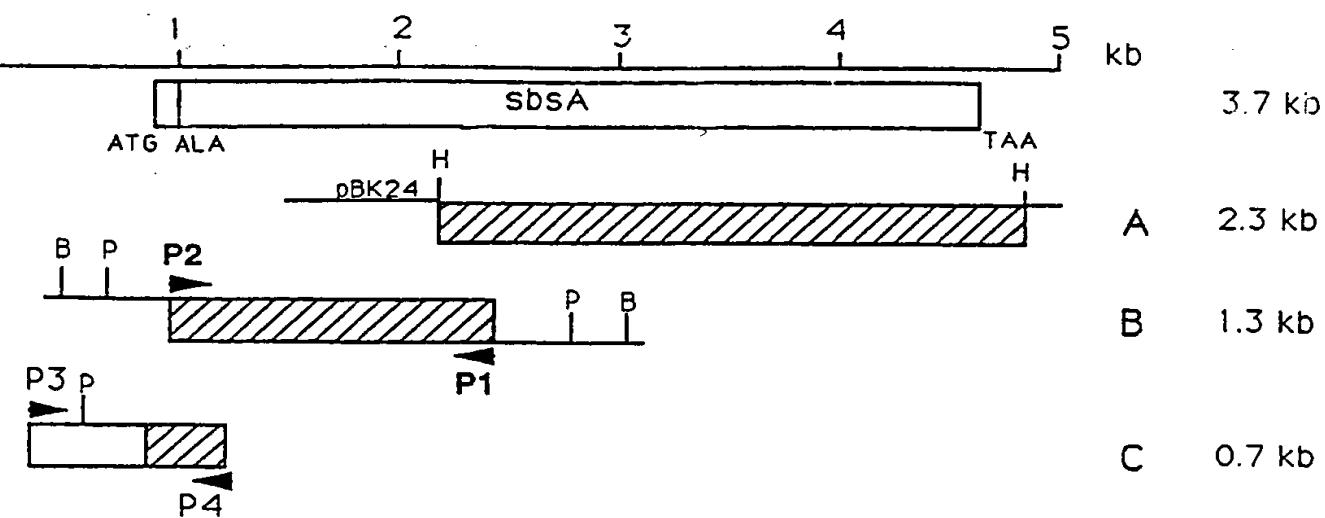


FIG. 1

- Leerseite -